SO HERE'S THE DEAL…

Regression Inference! What the heck does that mean? Basically, we're going back to our unit 02 content with the line of best fit (LSRL) and all of that goodness (less the transformations because that would make things too complicated!).

This is how it works:

We have a population of interest. We decide to see if there's some sort of correlation between two quantitative (measurable) variables.

Example: Population of interest = all ice cream cones
Variable 1 = number of grams of sugar in an ice cream cone
Variable 2 = number of calories in an ice cream cone

Is there a correlation between the number of grams of sugar in an ice cream cone and the number of calories in an ice cream cone?

Instead of using correlation in our STATE section, we're actually going to be measuring **slope**. We do this because the sign of r is the same as the sign of slope, and an r value of zero means a slope of zero. They're very closely related, but slope is a lot easier to calculate.

When we begin doing inference with this stuff, we have two options. We can either construct a confidence interval to estimate the true slope of the LSRL relating our two variables OR we can perform a significance test to determine if the slope is positive (> 0), negative (< 0), or non-zero ($\neq 0$).

We use β (capital beta) to represent the population slope and we use b to represent the sample slope (just like we use mu vs. x-bar or p vs. p-hat for means or proportions, respectively).

So, the next thing that happens, once we've chosen our population of interest and our two variables, is that we take a sample from the population. Let's say we choose sample size n = 50. That means we collect data on 50 ice cream cones and measure both the number of grams of sugar they contain as well as the number of calories they contain. We can plot all 50 points on a scatterplot and calculate the slope using our calculator
(STAT>CALC>8. LinReg a+bx). This gives us our sample slope b.

If we did this a bunch of times using all possible ice cream cone combinations with 50 cones, we could calculate a TON of sample slopes (b) and plot them, which would ideally result in a distribution that looks approximately Normal. We call this our sampling distribution (just like with means and proportions). The conditions we check to make sure this is the case are a little different than what we've done in the past. The acronym we use for the conditions is LINER and it's outlined pretty clearly in the notes. I have only seen like 2 FRAPPY problems EVER on an AP exam that require you to check these conditions. Most of the time, they say something like "assume the conditions required to perform inference are met." Just FYI. HOWEVER, they do appear more frequently in the MC section of AP exams.

> Let's go back to our single sample of 50 ice cream cones. We calculate the slope b (or maybe it's given to us in the problem). We are going to use this sample slope to create confidence interval that gives us plausible values for what the population slope might be. The equation is in your notes but follows the same pattern as before (point estimate +/- critical value* SE).

There are a couple of ways for you to get this information.

1. You are given the data and you calculate it by hand using the given equations.
2. You are given the data and you calculate it by calculator (STAT>TEST>G. LinRegTInt).
3. You are given these values in the problem (unlikely).
4. You are given a minitab output (see notes for where to get each value). **If this is the method you must use, YOU MUST SHOW THE PLUGGED IN VALUES IN THE EQUATION.**

If we are doing a significance test, we always assume (our null hypothesis) that the slope of the population regression line is 0. Our alternative is going to be either that it's non-zero, positive, or negative. The alternative should be stated in the problem somewhere and, again, the default significance level is 0.05.

> Let's go back to our ice cream problem. Our null is that $\beta = 0$. Let's say that the problem says something about someone assuming there is a positive relationship between number of grams of sugar and number of calories in an ice cream cone. That means our alternative will be $\beta > 0$. We do the significance test just like we have done in all of the other units.

When doing the DO section, you again have four ways to find the test statistic and p-value. They're the same as above. The calculator test is option F (LinRegTTest).

I feel like everything else is pretty clearly outlined in the notes BUT please send me an email or a message on Remind if you are confused about anything! Or, of course, ask your peers. ☺